

PENAKSIR RASIO REGRESI LINEAR UNTUK RATA-RATA POPULASI PADA SAMPLING ACAK SEDERHANA TANPA PENGEMBALIAN

Putri Susanti^{1*}, Arisman Adnan², Harison²

¹Mahasiswa Program S1 Matematika

²Dosen Jurusan Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Riau
Kampus Bina Widya Pekanbaru, 28293, Indonesia.

*putrisusanti92@yahoo.com

ABSTRACT

In this article we review three linear regression ratio estimators for the population mean on simple random sampling without replacement proposed by Singh *et al.* [*Statistics in Transition*, 10(2009): 85 – 100]. These estimators are biased estimators and their mean square error are determined. The estimator with the smallest mean square error is the most efficient one. We apply these estimators to show the efficiencies of examples.

Keywords: *Simple random sampling, ratio regression estimator, bias and mean square error*

ABSTRAK

Dalam artikel ini ditinjau ulang tiga penaksir rasio regresi linear untuk rata-rata populasi pada sampling acak sederhana tanpa pengembalian yang diajukan oleh Singh *et al.* [*Statistics in Transition*, 10(2009): 85 – 100]. Penaksir ini merupakan penaksir bias dan ditentukan *mean square error*. Penaksir dengan *mean square error* terkecil merupakan penaksir yang paling efisien. Penaksir ini digunakan untuk menunjukkan efisiensi dari contoh.

Kata Kunci: *Sampling acak sederhana, penaksir rasio regresi, bias dan mean square error*

1. PENDAHULUAN

Penaksir rasio regresi linear merupakan salah satu metode yang digunakan untuk meningkatkan ketelitian penaksir dengan mengambil manfaat hubungan antara dua variabel y_i dan x_i , dimana y_i adalah unit dari populasi berkarakter Y dan x_i adalah unit dari populasi berkarakter X . Penaksir rasio regresi linear menggunakan suatu variabel tambahan yang diketahui berkorelasi positif dengan variabel yang diteliti. Pada artikel ini informasi tambahan yang diketahui yaitu \bar{X} dan \bar{Z} yang merupakan rata-rata populasi dari variabel tambahan X dan Z . Variabel Y dan X , serta Y dan Z diharuskan berkorelasi positif sedangkan variabel X dan Z mungkin tidak berkorelasi.

Beberapa bentuk penaksir rasio penaksir regresi linear untuk rata-rata populasi telah dibahas dalam artikel sebelumnya antara lain oleh Srivenkataramana [4], Upadhyaya

dan Singh [5]. Pada artikel ini dibandingkan tiga penaksir rasio regresi untuk rata-rata populasi yang dibahas oleh Singh *et al* [3], yaitu

$$\hat{Y}_{RR}^{(1)} = \bar{y}_{lr} [(1 + w^*) - w^* u] \quad (1)$$

$$\hat{Y}_{RR}^{(2)} = \bar{y}_{lr} [(1 - w^{**}) + w^{**} u] \quad (2)$$

$$\hat{Y}_{RR}^{(3)} = \bar{y}_{lr} [(1 - \alpha) + \alpha u] \quad (3)$$

dengan $u = \frac{\bar{z}}{\bar{Z}}$, $w^* = \frac{\bar{Z}}{\alpha - \bar{Z}}$, $w^{**} = \frac{\bar{Z}}{\alpha + \bar{Z}}$, $\bar{y}_{lr} = \bar{y} + b_{yx}(\bar{X} - \bar{x})$ dan $0 < \alpha < 1$.

Ketiga penaksir rasio regresi linear tersebut merupakan penaksir yang bias. Ketelitian dari suatu penaksir yang bersifat bias ditinjau berdasarkan *mean square error* (MSE).

2. SAMPLING ACAK SEDERHANA

Sampling acak sederhana merupakan sebuah metode untuk mengambil n unit sampel dari N unit populasi dimana setiap unit mempunyai kesempatan yang sama untuk terambil menjadi anggota sampel. Pengambilan sampel dapat dilakukan dengan pengembalian atau tanpa pengembalian. Pada artikel ini pengambilan sampel dilakukan dengan sampling tanpa pengembalian. Untuk sampling tanpa pengembalian, peluang terambilnya anggota sampel berukuran n dari populasi berukuran N pada pengambilan yang pertama yaitu n/N . Pada pengambilan kedua peluang terambilnya sampel adalah $(n-1)/(N-1)$, dan seterusnya. Dapat disimpulkan bahwa peluang terambilnya seluruh n unit-unit tertentu yang terambil dalam n pengambilan adalah

$$\frac{n}{N} \cdot \frac{(n-1)}{(N-1)} \cdot \frac{(n-2)}{(N-2)} \cdots \frac{1}{(N-n+1)} = \frac{n!(N-n)!}{N!} = \frac{1}{C_n^N}.$$

Misalkan suatu populasi berukuran N dengan nilai variabel y_i untuk masing - masing unit $i = 1, 2, 3, \dots, N$, maka rata - rata populasi \bar{Y} adalah

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i.$$

Apabila diambil sampel berukuran n unit dengan nilai variabel y_i untuk masing-masing unit $i = 1, 2, 3, \dots, n$, maka rata-rata sampel \bar{y} adalah

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Teorema 2.1 [1: h. 27] Apabila sampel berukuran n diambil dari populasi berukuran N yang berkarakter Y , dengan sampling acak sederhana tanpa pengembalian maka variansi rata - rata sampel \bar{y} dinotasikan dengan $V(\bar{y})$ dan dirumuskan sebagai

$$V(\bar{y}) = \frac{1-f}{n} S_y^2$$

dengan $f = \frac{n}{N}$, $S_y^2 = \frac{\sum_{i=1}^N (y_i - \bar{Y})^2}{N-1}$ adalah variansi y_i pada populasi berkarakter Y .

Bukti: Dapat dilihat pada [1 : h. 27]. ■

Teorema 2.2 [1 : h. 29] Jika y_i, x_i adalah pasangan yang bervariasi pada unit dalam populasi dan \bar{y}, \bar{x} adalah rata-rata dari sampling acak sederhana yang diperoleh tanpa pengembalian berukuran n , maka kovariansinya adalah

$$Cov(\bar{X}, \bar{Y}) = \frac{1-f}{n} \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})(x_i - \bar{X}).$$

Bukti: Dapat dilihat pada [1 : h. 29]. ■

3. PENAKSIR REGRESI LINEAR UNTUK RATA-RATA POPULASI

Bentuk umum model regresi linear sederhana dinyatakan dalam persamaan

$$y_i = A + Bx_i + e_i \quad \text{untuk } i = 1, 2, 3, \dots, n. \quad (4)$$

Dari persamaan (4) maka kekeliruan ke- i ditulis

$$e_i = y_i - A - Bx_i \quad (5)$$

Dengan demikian jumlah kuadrat kekeliruan terhadap garis regresi ditulis

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - A - Bx_i)^2. \quad (6)$$

Dengan meminimumkan serta mengganti A dan B pada persamaan (6) dengan masing-masing taksiran a dan b , maka diperoleh

$$a = \bar{y} - b\bar{x}. \quad (7)$$

dan

$$b = \frac{S_{xy}}{S_x^2}. \quad (8)$$

Apabila garis regresi linear melalui titik pangkal $a = 0$, maka persamaan (7) menjadi

$$\bar{y} = b\bar{x}. \quad (9)$$

Dari persamaan (8), koefisien b yang diperoleh dari sampel digunakan juga untuk menaksir parameter populasi. Penaksir untuk rata-rata populasi dinotasikan dengan $\hat{\bar{Y}}$ dan dirumuskan dengan

$$\hat{\bar{Y}} = b\bar{X} \quad (10)$$

Selisih antara persamaan (10) dan persamaan (9) secara aljabar diperoleh

$$\hat{\bar{Y}} = \bar{y} + b(\bar{X} - \bar{x}).$$

Penaksir regresi linear untuk rata-rata populasi dinotasikan dengan $\hat{\bar{Y}}_{lr}$, yaitu

$$\hat{\bar{Y}}_{lr} = \bar{y} + b_{yx}(\bar{X} - \bar{x})$$

4. BIAS DAN MSE PENAKSIR RASIO REGRESI LINEAR UNTUK RATA-RATA POPULASI

Masing-masing penaksir yang dibahas merupakan penaksir bias, kemudian akan ditentukan bias dan MSE dari masing-masing penaksir. Untuk menentukan penaksir

yang lebih efisien dengan membandingkan MSE dari setiap penaksir. Suatu penaksir dikatakan lebih efisien apabila mempunyai MSE yang minimum.

Bias dan MSE penaksir rasio regresi linear untuk rata-rata populasi pada sampling acak sederhana tanpa pengembalian dijelaskan untuk setiap penaksir berikut.

Besarnya bias penaksir pertama pada persamaan (1) adalah

$$B(\hat{Y}_{RR}^{(1)}) = \frac{\bar{Z}}{\alpha - \bar{Z}} \bar{Y} - \frac{1}{\alpha - \bar{Z}} (Cov(\bar{y}_{lr}, \bar{z}) + \bar{Y} \bar{Z})$$

dan MSE -nya adalah

$$MSE(\hat{Y}_{RR}^{(1)}) \approx \frac{1-f}{n} \bar{Y}^2 (C_y^2 - \rho_{yx}^2 C_y^2 + (w^*)^2 C_z^2 + 2(w^*) \rho_{yx} \rho_{xz} C_y C_z - 2(w^*) \rho_{yz} C_y C_z).$$

Besarnya bias penaksir kedua pada persamaan (2) yaitu

$$B(\hat{Y}_{RR}^{(2)}) = -\frac{\bar{Z}}{\alpha + \bar{Z}} \bar{Y} + \frac{1}{\alpha + \bar{Z}} (Cov(\bar{y}_{lr}, \bar{z}) + \bar{Y} \bar{Z})$$

dan MSE -nya adalah

$$MSE(\hat{Y}_{RR}^{(2)}) \approx \frac{1-f}{n} \bar{Y}^2 (C_y^2 - \rho_{yx}^2 C_y^2 + (w^{**})^2 C_z^2 - 2(w^{**}) \rho_{yx} \rho_{xz} C_y C_z + 2(w^{**}) \rho_{yz} C_y C_z).$$

Besarnya bias penaksir ketiga pada persamaan (3) yaitu

$$B(\hat{Y}_{RR}^{(3)}) = -\alpha \bar{Y} + \frac{\alpha}{\bar{Z}} (Cov(\bar{y}_{lr}, \bar{z}) + \bar{Y} \bar{Z})$$

dan MSE -nya adalah

$$MSE(\hat{Y}_{RR}^{(3)}) \approx \frac{1-f}{n} \bar{Y}^2 (C_y^2 - \rho_{yx}^2 C_y^2 + (\alpha)^2 C_z^2 - 2(\alpha) \rho_{yx} \rho_{xz} C_y C_z + 2(\alpha) \rho_{yz} C_y C_z).$$

Berikut ini dibandingkan MSE masing-masing penaksir, yaitu

1. Perbandingan antara $MSE(\hat{Y}_{RR}^{(1)})$ dengan $MSE(\hat{Y}_{RR}^{(2)})$, diperoleh

$$MSE(\hat{Y}_{RR}^{(2)}) < MSE(\hat{Y}_{RR}^{(1)}),$$

jika

$$\rho_{yz} > \frac{C_z (w^* - w^{**}) + 2\rho_{yx} \rho_{xz} C_y}{2C_y}.$$

2. Perbandingan antara $MSE(\hat{Y}_{RR}^{(1)})$ dengan $MSE(\hat{Y}_{RR}^{(3)})$, diperoleh

$$MSE(\hat{Y}_{RR}^{(3)}) < MSE(\hat{Y}_{RR}^{(1)}),$$

jika

$$\rho_{yz} > \frac{C_z (w^* - \alpha) + 2\rho_{yx} \rho_{xz} C_y}{2C_y}.$$

3. Perbandingan antara $MSE(\hat{Y}_{RR}^{(2)})$ dengan $MSE(\hat{Y}_{RR}^{(3)})$, diperoleh

$$MSE(\hat{Y}_{RR}^{(3)}) < MSE(\hat{Y}_{RR}^{(2)}),$$

jika

$$\rho_{yz} > \frac{-(C_z (w^{**} + \alpha) - 2\rho_{yx} \rho_{xz} C_y)}{2C_y}.$$

Contoh Perhatikan data pada Tabel 1 mengenai permintaan masyarakat terhadap air minum isi ulang di kota Pekanbaru [2] dengan menggunakan variabel tambahan yaitu pendapatan masyarakat X dan jumlah anggota rumah tangga Z . Dengan menganggap data tersebut populasi, akan ditentukan penaksir rasio regresi linear yang lebih efisien untuk menaksir rata-rata permintaan air minum isi ulang yakni dengan menggunakan syarat penaksir lebih efisien yang diperoleh sebelumnya. Secara umum untuk menentukan penaksir rasio regresi linear yang lebih efisien dapat ditunjukkan dengan menghitung nilai MSE dari masing-masing penaksir yang diberikan.

Tabel 1. Permintaan masyarakat terhadap air minum isi ulang di Kota Pekanbaru, Riau

No	Permintaan Air Minum Isi Ulang (Galon) (Y)	Pendapatan Masyarakat (Rp/Bulan) (X)	Jumlah Anggota Rumah Tangga (Orang) (Z)	No	Permintaan Air Minum Isi Ulang (Galon) (Y)	Pendapatan Masyarakat (Rp/Bulan) (X)	Jumlah Anggota Rumah Tangga (Orang) (Z)
1	5	1.800.000	3	46	5	2.150.000	4
2	6	2.300.000	3	47	7	3.500.000	4
3	7	3.000.000	3	48	11	2.450.000	6
4	11	2.750.000	5	49	7	1.850.000	4
5	4	1.500.000	1	50	4	2.150.000	3
6	7	2.800.000	3	51	5	2.000.000	3
7	8	2.750.000	4	52	6	1.850.000	4
8	7	4.800.000	3	53	5	2.400.000	3
9	8	1.750.000	5	54	6	2.250.000	4
10	4	1.200.000	1	55	9	2.500.000	5
11	4	3.200.000	6	56	12	4.500.000	6
12	8	2.000.000	4	57	10	3.200.000	5
13	4	1.500.000	1	58	10	3.000.000	5
14	3	1.600.000	1	59	12	4.400.000	6
15	8	2.000.000	5	60	8	3.500.000	4
16	9	2.500.000	4	61	7	2.250.000	4
17	13	2.850.000	6	62	4	1.800.000	3
18	9	3.000.000	5	63	7	3.500.000	4
19	7	2.200.000	4	64	6	2.200.000	4
20	4	1.500.000	1	65	11	3.200.000	6
21	10	2.500.000	5	66	10	3.500.000	5
22	8	2.500.000	4	67	9	2.750.000	5
23	13	3.500.000	7	68	10	2.500.000	6
24	6	2.500.000	3	69	10	3.500.000	5
25	7	2.300.000	4	70	5	2.400.000	3
26	10	3.500.000	5	71	9	2.800.000	4
27	12	5.400.000	6	72	9	3.500.000	5
28	5	2.200.000	2	73	12	3.400.000	6

29	5	2.500.000	3	74	10	2.850.000	5
30	9	2.800.000	5	75	4	2.400.000	1
31	11	3.500.000	5	76	7	2.300.000	4
32	7	2.550.000	4	77	10	3.200.000	5
33	4	1.550.000	1	78	8	2.560.000	4
34	5	1.900.000	3	79	11	2.800.000	5
35	8	2.000.000	4	80	11	3.500.000	5
36	9	3.800.000	5	81	8	3.000.000	4
37	6	2.800.000	3	82	9	3.500.000	5
38	7	2.150.000	4	83	4	1.850.000	1
39	7	2.300.000	4	84	10	2.500.000	5
40	8	3.500.000	4	85	6	3.200.000	6
41	14	2.950.000	8	86	7	2.700.000	4
42	10	3.500.000	5	87	3	1.200.000	1
43	11	2.650.000	6	88	7	3.500.000	4
44	6	3.000.000	3	89	13	5.300.000	6
45	5	2.730.000	3	-	-	-	-

Sumber [2].

Dari data tersebut selanjutnya dihitung beberapa nilai-nilai yang diperlukan untuk membandingkan *MSE* dari ketiga penaksir seperti pada Tabel 2.

Tabel 2. Nilai parameter dari populasi

N	89	n	30
\bar{Y}	7,786	S_{yx}	1.419.923,351
\bar{X}	2.729.662,921	S_{yz}	3,429
\bar{Z}	4,101	S_{xz}	739.878,921
S_x	2.864.539,027	ρ_{yx}	0,186
S_y	2,665	ρ_{yz}	0,862
S_z	1,493	ρ_{xz}	0,173
C_y	0,342	C_x	1,049
C_z	0,354	α	0,5

Untuk mengetahui seberapa dekat nilai taksiran terhadap parameter populasi, dapat dilakukan dengan menghitung nilai pada penaksir $\hat{Y}_{RR}^{(1)}$, $\hat{Y}_{RR}^{(2)}$, dan $\hat{Y}_{RR}^{(3)}$, seperti yang dicantumkan pada Tabel 3.

Tabel 3. Nilai *MSE* dan nilai taksiran dari masing-masing penaksir

Penaksir	<i>MSE</i>	Nilai taksiran
$\hat{Y}_{RR}^{(1)}$	0,675	6,479
$\hat{Y}_{RR}^{(2)}$	0,523	6,639
$\hat{Y}_{RR}^{(3)}$	0,327	6,895

Dari Tabel 3 dapat dilihat bahwa $MSE\left(\hat{Y}_{RR}^{(3)}\right) < MSE\left(\hat{Y}_{RR}^{(2)}\right) < MSE\left(\hat{Y}_{RR}^{(1)}\right)$, dengan demikian penaksir $\hat{Y}_{RR}^{(3)}$ adalah penaksir yang paling efisien.

5. KESIMPULAN

Berdasarkan pembahasan yang telah dikemukakan pada artikel ini, maka dapat disimpulkan bahwa penaksir rasio regresi linear $\hat{Y}_{RR}^{(3)}$ lebih efisien dari penaksir rasio regresi linear $\hat{Y}_{RR}^{(1)}$ dan $\hat{Y}_{RR}^{(2)}$.

DAFTAR PUSTAKA

- [1] Cochran, W.G. 1991. *Teknik Penarikan Sampel, Edisi Ketiga*. Terj. dari *Sampling Techniques*, oleh Rudiansyah & E.R Osman. Penerbit Universitas Indonesia, Jakarta.
- [2] Sihite, Wenny. 2012. Analisis Permintaan Masyarakat Terhadap Air Minum Oksigen Isi Ulang (*New Water Oxygen Hexagonal*) di Kota Pekanbaru. Skripsi, Fakultas Ekonomi Universitas Riau, Pekanbaru.
- [3] Singh, H.P., L.N. Upadhyaya, & Premchandra. 2009. An Improved Version of Regression Ratio Estimator With Two Auxiliary Variables in Sample Surveys, *Statistics In Transition*. 10(1): 85-100.
- [4] Srivenkataramana, T. 1980. A Dual to Ratio Estimator in a Sample Surveys, *Biometrika*. 67(1): 199-204.
- [5] Upadhyaya, L.N. & H.P. Singh. 1984. Combination of Regression and Ratio Estimate, *Current Science*. 53(6): 294-296.